



Crimes in Boston (2018)

Nicholas Howard and Alyssa Volivar



Dataset Overview

- Collection of crime incident reports provided by the Boston Police department from 2015-2018
- Dataset was modified by Kaggle users
 - Year, Month, Day, Hour of Day, Latitude, Longitude variables added
- Original dataset contains 260,760 rows and 17 variables
- Truncated dataset we chose to work with contained only 2018 incident reports, which consists of 74,356 rows and 17 variables

Description of Variables	
Variable	Description
INCIDENT_NUMBER:	Internal BPD report number
OFFENSE_CODE:	Numerical code of offense description
OFFENSE_CODE_GROUP:	Internal categorization of [offense_description]
OFFENSE_DESCRIPTION:	Primary descriptor of incident
DISTRICT:	What district the crime was reported in
REPORTING_AREA:	RA number associated with the where the crime was reported from.
SHOOTING:	Indicated a shooting took place.
OCCURRED_ON_DATE:	Earliest date and time the incident could have taken place
YEAR:	Year the incident took place
MONTH:	Month the incident took place
DAY_OF_WEEK:	Day of the week the incident took place
HOUR:	Hour the incident took place
UCR_PART:	Universal Crime Reporting Part number (1,2, 3)
STREET:	Street name the incident took place
LATITUDE:	Parallel the incident occurred in degrees (40.0 = 40° N of the equator)
LONGITUDE:	Meridian the incident occurred (-70.0 = 70° west of the Greenwich Meridian)
LOCATION:	Geolocation (Lat, Long), Reference: Boston = 42.35866° N, -71.05674° E



Problem and Research Questions

Crime rates can affect quality of living for residents, and can also impact tourism, the willingness of a business to expand operations in the city, discourage those looking to relocate to Boston, etc.

- Can we predict the type of offense that will occur?
- How likely is a shooting or a violent crime likely to occur?
- What relationships are there between the number of offenses, location, and time?
- What relationships are there between the type of offenses, location, and time?



Variable: District

- Contained 560 missing values; rows with missing values excluded

Table 4: Frequency of Crime by District

District Code	District Name	Frequency ¹	Population ²	Frequency/Population
B2	Roxbury	11380	48454	0.235
D4	South End	9799	24577	0.399
C11	Dorchester	9487	114235	0.083
A1	Downtown	8439	11215	0.752
B3	Mattapan	8385	22600	0.371
C6	South Boston	5877	33311	0.176
D14	Brighton	4591	45801	0.100
E18	Hyde Park	4245	30637	0.139
E13	Jamaica Plain	4103	37468	0.110
E5	West Roxbury	3056	30446	0.100
A7	East Boston	2953	40508	0.073
A15	Charlestown	1481	16439	0.090

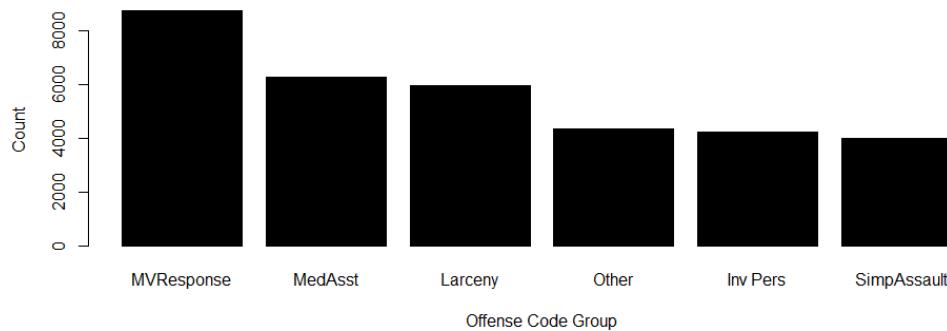


Variable: Offense Groups

Table 5: Top 6 Offense Codes in 2018

Offense Code Group	Frequency	Relative Frequency
Motor Vehicle Accident Response	8724	12%
Medical Assistance	6292	8%
Larceny	5949	8%
Other	4372	6%
Investigate Person	4241	6%
Simple Assault	4013	5%

Top 6 Event Code Group in 2018

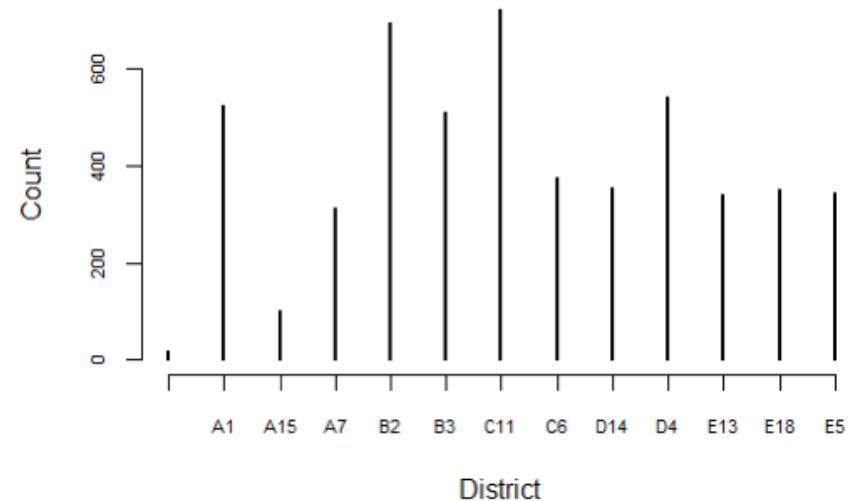


Variable: Offense Codes and Descriptions

Table 6: Top 6 Offense Code Descriptions

Code	Frequency	Description
3006	5203	SICK/INJURED/MEDICAL - PERSON
3115	4241	INVESTIGATE PERSON
3831	3790	M/V - LEAVING SCENE - PROPERTY DAMAGE
802	3694	ASSAULT SIMPLE - BATTERY
3301	3433	VERBAL DISPUTE
1402	3155	VANDALISM

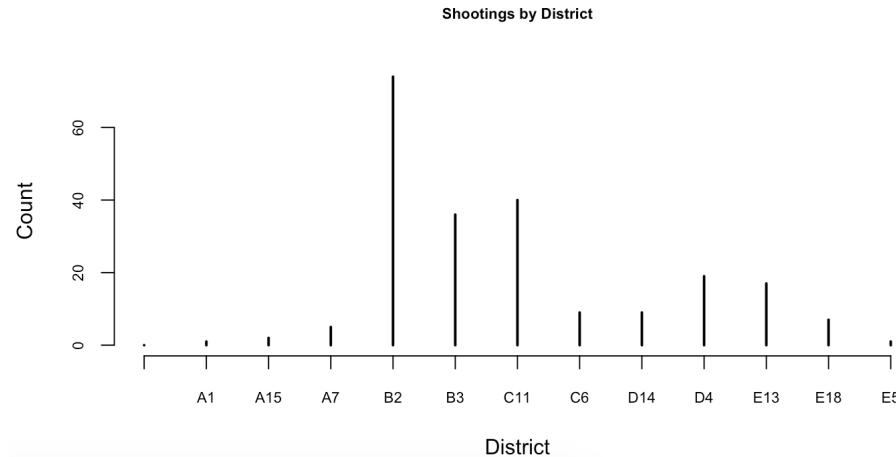
Offense Code 3006 by District



Variable: Shooting

Table 5: Shootings by Frequency

Offense Code Description	Count
ASSAULT - AGGRAVATED - BATTERY	92
MURDER, NON-NEGLIGENT MANSLAUGHTER	28
WARRANT ARREST	23
WEAPON - FIREARM - CARRYING / POSSESSING, ETC	15
BALLISTICS EVIDENCE/FOUND	11
SEARCH WARRANT	8
WEAPON - FIREARM - OTHER VIOLATION	6



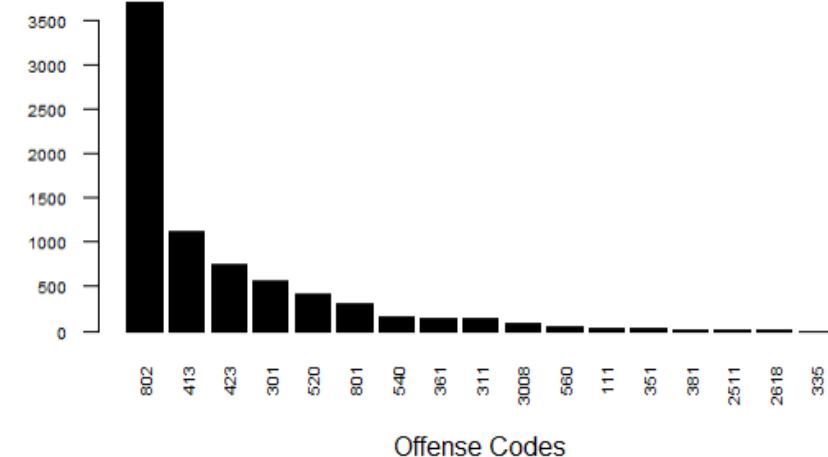
New Variable: Violent Crimes

- Created new binary variable using ifelse() to differentiate between violent and non-violent crimes

Table 6: Violent Crime Offense Code Descriptions by Frequency

Offense Code	Offense Code Description	Count
802	ASSAULT SIMPLE - BATTERY	3694
413	ASSAULT - AGGRAVATED - BATTERY	1125
423	ASSAULT - AGGRAVATED	752
301	ROBBERY - STREET	558
520	BURGLARY - RESIDENTIAL - FORCE	425
801	ASSAULT - SIMPLE	306
540	BURGLARY - COMMERCIAL - FORCE	156
361	ROBBERY - OTHER	146
311	ROBBERY - COMMERCIAL	141
3008	SUICIDE / SUICIDE ATTEMPT	86
560	BURGLARY - OTHER - FORCE	53
111	MURDER, NON-NEGLIGENCE MANSLAUGHTER	37
351	ROBBERY - BANK	26
381	ROBBERY - CAR-JACKING	17
2511	#N/A	10
2618	EXPLOSIVES - POSSESSION OR USE	4
	ROBBERY - UNARMED - CHAIN STORE	1

Violent Crime in 2018



335

Model: Violent Crimes

2.1 Model 1: Violent Crime predicted by Offense Code

- Logistic Regression
- Attempt to predict the occurrence of a violent crime based on the Offense Code
- **AIC:** 27619.

$$\log \left(\frac{\hat{\theta}(\mathbf{x})}{1-\hat{\theta}(\mathbf{x})} \right) = 1.419364576 + (-0.003020421_{\text{OFFENSE_CODE}})$$

	Coefficient	Pr(> z)
Intercept	1.419364576	<2e-16*
OFFENSE_CODE	-0.003020421	<2e-16*



Model: Violent Crimes

2.1 Model 1: Violent Crime predicted by Offense Code

```
call:  
glm(formula = crime2018$violentcrime ~ crime2018$OFFENSE_CODE,  
     family = binomial(link = "logit"))  
  
Deviance Residuals:  
    Min      1Q      Median      3Q      Max  
-1.1126 -0.1771 -0.0261 -0.0167  3.9157  
  
Coefficients:  
              Estimate Std. Error z value Pr(>|z|)  
(Intercept) 1.419e+00 3.704e-02 38.32 <2e-16 ***  
crime2018$OFFENSE_CODE -3.020e-03 4.833e-05 -62.50 <2e-16 ***  
---  
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
Null deviance: 48879  on 74355  degrees of freedom  
Residual deviance: 27615  on 74354  degrees of freedom  
AIC: 27619  
  
Number of Fisher Scoring iterations: 9
```



Model: Violent Crimes

2.1 Model 2: Violent Crime predicted by District

- Logistic Regression
- Attempt to predict the occurrence of a violent crime based on the District in which the crime occurred.
- **AIC:** 48725.

$$\log\left(\frac{\hat{\theta}(\mathbf{x})}{1-\hat{\theta}(\mathbf{x})}\right) = -2.8716796 + 0.9484626_{A1} + 0.6585997_{A15} + 0.7884227_{A7} + 0.7766229_{B2} + 0.7308135_{B3} + 0.7275425_{C11} + 0.9484626_{A1} + 0.6010544_{C6} + 0.3457157_{D14} + 0.7081290_{D4}$$

	Coefficient	Pr(> z)		Coefficient	Pr(> z)
Intercept	-2.8716796	< 2e-16*	C11	0.7275425	0.000135*
A1	0.9484626	6.38e-07*	C6	0.6010544	0.001837*
A15	0.6585997	0.001458*	D14	0.3457157	0.077673
A7	0.7884227	6.07e-05*	D4	0.7081290	0.000203
B2	0.7766229	4.38e-05*	E13	0.5422600	0.005550*
B3	0.7308135	0.000130*	E18	0.3783211	0.054028
			E5	0.5254899	0.008049*



Model: Violent Crimes

2.1 Model 2: Violent Crime predicted by District

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.5223	-0.4818	-0.4664	-0.4276	2.4194

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.8717	0.1877	-15.302	< 2e-16 ***
crime2018\$DISTRICTA1	0.9485	0.1905	4.979	6.38e-07 ***
crime2018\$DISTRICTA15	0.6586	0.2069	3.183	0.001458 **
crime2018\$DISTRICTA7	0.7884	0.1966	4.010	6.07e-05 ***
crime2018\$DISTRICTB2	0.7766	0.1900	4.086	4.38e-05 ***
crime2018\$DISTRICTB3	0.7308	0.1910	3.826	0.000130 ***
crime2018\$DISTRICTC11	0.7275	0.1906	3.816	0.000135 ***
crime2018\$DISTRICTC6	0.6011	0.1929	3.115	0.001837 **
crime2018\$DISTRICTD14	0.3457	0.1959	1.764	0.077673 .
crime2018\$DISTRICTD4	0.7081	0.1906	3.716	0.000203 ***
crime2018\$DISTRICTE13	0.5423	0.1955	2.773	0.005550 **
crime2018\$DISTRICTE18	0.3783	0.1964	1.927	0.054028 .
crime2018\$DISTRICTE5	0.5255	0.1983	2.650	0.008049 **

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 48879 on 74355 degrees of freedom

Residual deviance: 48699 on 74343 degrees of freedom

AIC: 48725

Number of Fisher Scoring iterations: 5



Model Selection

2.1 Model 2: Violent Crime predicted by District

- Logistic Regression
- Attempt to predict the occurrence of a violent crime based on the District in which the crime occurred
- **AIC:** 48725.

**We used this model to predict probabilities
of the occurrence of violent crime
based on the district in which the
crime occurred**

District	Average of Predicted Probability
A1	0.10160162
A15	0.101441264
A7	0.101529551
B2	0.101551521
B3	0.101685976
C11	0.101605477
C6	0.101797161
D14	0.101987732
D4	0.101769009
E13	0.101951709
E18	0.101090801
E5	0.10148344



Model Selection

2.1 Model 2: Violent Crime predicted by District

District	Count
A1	1076
A15	146
A7	327
B2	1247
B3	882
C11	995
C6	550
D14	340
D4	1010
E13	364
E18	324
E5	267

District	Average of Predicted Probability
A1	0.10160162
A15	0.101441264
A7	0.101529551
B2	0.101551521
B3	0.101685976
C11	0.101605477
C6	0.101797161
D14	0.101987732
D4	0.101769009
E13	0.101951709
E18	0.101090801
E5	0.10148344



Conclusion and Future Works

- Limitations of this dataset: few numerical values to work with, variables mostly relate to time, location, and type of offense
- In future studies it would be helpful to have variables such as income per district, types of households, distance to transit, education levels, and other economic and social variables that might influence crime rates



Sources

- Antolos, D., Liu, D., Ludu, A., & Vincenzi, D. (2013). Burglary Crime Analysis Using Logistic Regression. In S. Yamamoto (Eds.), *Human Interface and the Management of Information. Information and Interaction for Learning, Culture, Collaboration and Business* (pp. 549-558). London: Springer.
- Boston Police Department. (n.d.). Districts. Retrieved from <https://bpdnews.squarespace.com/districts/>
- Boston Redevelopment Authority Research Division. *Boston in Context: Neighborhoods* [PDF Document]. Retrieved from bostonplans.org: <http://www.bostonplans.org/getattachment/e50e791c-caa4-41b0-9d14-35f88cff77af>
- Boston.gov (2015). Crime Incident Reports (August 2015-To Date) (Source: New System). Retrieved from <https://data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system>
- Federal Bureau of Investigation (n.d.). Uniform Crime Reporting (UCR) Program. Retrieved from <https://www.fbi.gov/services/cjis/ucr>
- Federal Bureau of Investigation (n.d.). Uniform Crime Report: Crime in the United States, 2011 . Retrieved from <https://ucr.fbi.gov/crime-in-the-u.s/2011/crime-in-the-u.s.-2011/offense-definitions>
- Kaggle.com (2018). *Crimes in Boston* [Data file]. Retrieved from <https://www.kaggle.com/ankkur13/boston-crime-data/activity>
- Nolan, J. (2004). Establishing the statistical relationship between population size and UCR crime rate: its impact and implications. *Journal of Criminal Justice*, 32(6), 547-555. doi:10.1016/j.jcrimjus.2004.08.002
- Reiss, J. (2019, April 22). How population numbers in Boston and Massachusetts changed over the past 8 years. *The Boston Globe*.

